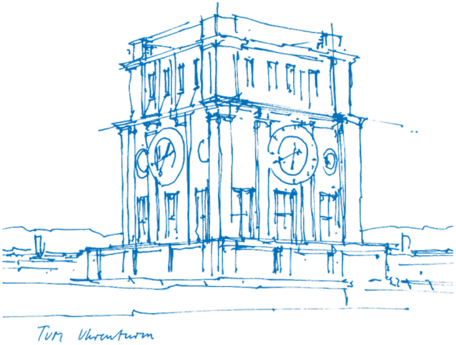
Munich Center for Machine Learning (MCML)
TUM School of Computation, Information and Technology
Technical University of Munich

Confidence in Causal Inference under Structure Uncertainty

David Strieder, Mathias Drton

European Causal Inference Meeting
Copenhagen, Denmark, April 17-19, 2024



Starting point

- **Research question:** What is the total causal effect of X_i on X_j? Confidence?
- **Given:** Observational data in form of n samples of $(X_1, ..., X_d)$.
- **Problem:** Causal structure unknown.

Starting point

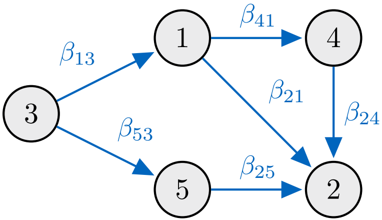
- **Research question:** What is the total causal effect of X_i on X_j? Confidence?
- **Given:** Observational data in form of n samples of $(X_1, ..., X_d)$.
- **Problem:** Causal structure unknown.
- Naive two-step approach?
  (1) Learn causal structure.
  (2) Calculate confidence intervals for causal effects in inferred model.

Setup

Underlying Linear SCM with equal error variances

- **Example:**

$$
\begin{aligned}
X_1 &= \beta_{13} X_3 + \varepsilon_1 \\
X_2 &= \beta_{21} X_1 + \beta_{24} X_4 + \beta_{25} X_5 + \varepsilon_2 \\
X_3 &= \varepsilon_3 \\
X_4 &= \beta_{41} X_1 + \varepsilon_4 \\
X_5 &= \beta_{25} X_3 + \varepsilon_5
\end{aligned}
$$



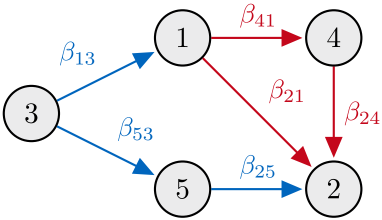
where $\varepsilon_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$

Setup

Underlying Linear SCM with equal error variances

- **Example:** Causal effect $\mathcal{C}(1 \to 2) := \frac{\mathrm{d}}{\mathrm{d}x_1}\mathbb{E}[X_2 | \operatorname{do}(X_1 = x_1)] = \Sigma_{12|p(1)}/\Sigma_{11|p(1)}$.

$$
\begin{aligned}
X_1 &= \beta_{13}X_3 + \varepsilon_1 \\
X_2 &= \beta_{21}X_1 + \beta_{24}X_4 + \beta_{25}X_5 + \varepsilon_2 \\
X_3 &= \varepsilon_3 \\
X_4 &= \beta_{41}X_1 + \varepsilon_4 \\
X_5 &= \beta_{25}X_3 + \varepsilon_5
\end{aligned}
$$



where $\varepsilon_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$

Setup

- **Main Idea:** Use test inversion.

Tests for $\mathrm{H}_0^{(\psi)} : \mathcal{C}(i \to j) = \psi$ $\longrightarrow$ Confidence interval for $\mathcal{C}(i \to j)$

- **Goal:** Construct suitable **tests for all possible effects.**

Setup

- **Main Idea:** Use test inversion.

Tests for $\mathsf{H}_0^{(\psi)} : \mathcal{C}(i \to j) = \psi$	$\longrightarrow$	Confidence interval for $\mathcal{C}(i \to j)$

- **Goal:** Construct suitable **tests for all possible effects.**

- **Difficulty:** Each Hypothesis of fixed effect is **union of single hypotheses** over all DAGs on d nodes.

$$\mathsf{H}_0^{(\psi)} := \bigcup_{G \in \mathcal{G}(d)} \mathsf{H}_0^{(\psi)}(G)$$

Setup

- Linear SCM with equal error variances: $\mathcal{M} := \bigcup_{G \in \mathcal{G}(d)} \mathcal{M}(G)$, where

$$\mathcal{M}(G) = \left\{ \Sigma \in \mathsf{PD}(d) : \exists \sigma^2 > 0 \text{ with } \sigma^2 = \Sigma_{k,k|p(k)} \quad \forall\, k = 1, \ldots, d \right\}.$$

Setup

- Linear SCM with equal error variances: $\mathcal{M} := \bigcup_{G\in\mathcal{G}(d)} \mathcal{M}(G)$, where

$$\mathcal{M}(G) = \left\{\Sigma \in \mathsf{PD}(d) : \exists\sigma^2 > 0 \text{ with } \sigma^2 = \Sigma_{k,k|p(k)} \quad \forall\, k = 1,\dots,d\right\}.$$

- Hypothesis of fixed total causal effect: $\mathcal{M}_\psi := \bigcup_{G\in\mathcal{G}(d)} \mathcal{M}_\psi(G)$, where

$$\mathcal{M}_\psi(G) := \left\{\Sigma \in \mathcal{M}(G) : \psi = \Sigma_{j,i|p(i)}/\Sigma_{i,i|p(i)}\right\}.$$

Setup

- Linear SCM with equal error variances: $\mathcal{M} := \bigcup_{G \in \mathcal{G}(d)} \mathcal{M}(G)$, where

$$\mathcal{M}(G) = \left\{ \Sigma \in \mathrm{PD}(d) : \exists \sigma^2 > 0 \text{ with } \sigma^2 = \Sigma_{k,k|p(k)} \quad \forall\, k = 1, \ldots, d \right\}.$$

- Hypothesis of fixed total causal effect: $\mathcal{M}_\psi := \bigcup_{G \in \mathcal{G}(d)} \mathcal{M}_\psi(G)$, where

$$\mathcal{M}_\psi(G) := \left\{ \Sigma \in \mathcal{M}(G) : \psi = \Sigma_{j,i|p(i)} / \Sigma_{i,i|p(i)} \right\}.$$

- **Task**: For all $\psi \in \mathbb{R}$ and $G \in \mathcal{G}(d)$ invert joint test of structure and effect size:

$$\mathrm{H}_0^{(\psi)}(G) : \Sigma \in \mathcal{M}_\psi(G) \quad \text{against} \quad \mathrm{H}_1 : \Sigma \in \mathcal{M} \setminus \mathcal{M}_\psi(G).$$

Maximum Likelihood Estimation

Munich Center for Machine Learning

- **Problem**: Maximizing the Gaussian likelihood

$$\frac{2}{n}\ell_n(\Sigma) = -\log\det(2\pi\Sigma) - \mathrm{tr}(\Sigma^{-1}\hat{\Sigma})$$

with $\Sigma \in \mathcal{M}_\psi(G)$ is equivalent to minimizing

$$\mathrm{tr}((I_d - B)^T(I_d - B)\hat{\Sigma}))$$

where $B \in \mathbb{R}^G$ represents the direct causal effects between variables.

Maximum Likelihood Estimation

- **Problem**: Maximizing the Gaussian likelihood

$$\frac{2}{n}\ell_n(\Sigma) = -\log\det(2\pi\Sigma) - \mathrm{tr}(\Sigma^{-1}\hat{\Sigma})$$

with $\Sigma \in \mathcal{M}_\psi(G)$ is equivalent to minimizing

$$\mathrm{tr}((I_d - B)^T(I_d - B)\hat{\Sigma}))$$

where $B \in \mathbb{R}^G$ represents the direct causal effects between variables.

- Fixing total causal effects is complex **polynomial constraints** on direct effects, namely

$$(I_d - B)^{-1}_{j,i} = \psi.$$

Dual Maximum Likelihood Estimation

- **Solution**: Maximizing the Dual likelihood

$$\frac{2}{n}\ell_n^{dual}(\Sigma) := -\log\det(2\pi\Sigma^{-1}) - \mathrm{tr}(\Sigma\widehat{\Sigma}^{-1})$$

with $\Sigma \in \mathcal{M}_\psi(G)$ is equivalent to minimizing

$$\mathrm{tr}((I_d - B)^{-1}(I_d - B)^{-T}\hat{\Sigma})) = \mathrm{tr}((I_d - T)^T(I_d - T)\hat{\Sigma}))$$

where $T \in \mathbb{R}^{-G}$ represents the negative total causal effects between variables.

Dual Maximum Likelihood Estimation

- **Solution**: Maximizing the Dual likelihood

$$\frac{2}{n}\ell_n^{dual}(\Sigma) := -\log\det(2\pi\Sigma^{-1}) - \mathrm{tr}(\Sigma\widehat{\Sigma}^{-1})$$

  with $\Sigma \in \mathcal{M}_\psi(G)$ is equivalent to minimizing

$$\mathrm{tr}((I_d - B)^{-1}(I_d - B)^{-T}\hat{\Sigma})) = \mathrm{tr}((I_d - T)^T(I_d - T)\hat{\Sigma}))$$

  where $T \in \mathbb{R}^{-G}$ represents the negative total causal effects between variables.

- Fixed total effect constraint only pertains to one parameter, namely

$$T_{i,j} = -\psi$$

Main Result

- **Main steps**:

  (1) Intersection union test.

  (2) Stochastic upper bound by relaxing alternative.

  (3) Dual-LRT with conservative critical values from upper bound.

Main Result

- **Main steps**:

  (1) Intersection union test.

  (2) Stochastic upper bound by relaxing alternative.

  (3) Dual-LRT with conservative critical values from upper bound.

- **Result:** Asymptotic $(1-\alpha)$-confidence set for causal effect $\mathcal{C}(i \to j)$ is

$$\{\psi \in \mathbb{R} : \min_{G \in \mathcal{G}(d)\,:\, i <_G j} \text{dual-}\lambda_n^{(\psi)}(G) \leq \chi^2_{d,1-\alpha}\} \cup \{0 : \min_{G \in \mathcal{G}(d)\,:\, j <_G i} \text{dual-}\lambda_n^{(0)}(G) \leq \chi^2_{d-1,1-\alpha}\}$$

Computation

- **Bottleneck:** Superexponential growth of possible causal structures with nodes.

Computation

- **Bottleneck:** Superexponential growth of possible causal structures with nodes.
- 'Only' need to consider complete DAGs: d factorial structures.
- Branch and bound type search algorithm through causal orderings. Feasible up to 12 involved variables (already uncertainty over more than 10^{26} structures).

Conclusion

- Closed-form solution for constructing confidence regions for total causal effects that:
  - account for **causal structure uncertainty**
  - as well as **statistical uncertainty** about the numerical size of the effect.
- Conceptual ideas of leveraging test inversions of joint tests for causal structure and effect size generalizable to other modeling assumptions.
- Matrix inversion interplay between direct and total effects that is behind the use of dual likelihood can be exploited in other causal inference tasks.

Thank you!

Munich Center for Machine Learning

- Our related papers:
  - Strieder and Drton (2024). *Dual Likelihood for Causal Inference under Structure Uncertainty*. CLeaR 24. Preprint at arXiv.
  - Strieder and Drton (2023). *Confidence in Causal Inference under Structure Uncertainty in Linear Causal Models with Equal Variances*. J. Causal Inference 11 (1).
  - Strieder, Freidling, Haffner and Drton (2021). *Confidence in Causal Discovery with Linear Causal Models*. UAI 21.